RL-TR-97-73
In-House Report
August 1997

A NEW EFFICIENT ALGORITHM FOR PDF APPROXIMATION

Lisa K. Slaski and Murali Rangaswamy

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

19971007 215

DTIC QUALITY INSPECTED 4

Rome Laboratory
Air Force Materiel Command
Rome, New York

This report has been reviewed by the Rome Laboratory Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.
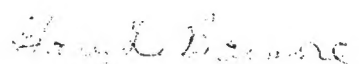
RL-TR-97-73 has been reviewed and is approved for publication.

APPROVED:

JAMES W. CUSACK, Chief
Surveillance Division
Surveillance & Photonics Directorate

FOR THE DIRECTOR:

GARY D. BARMORE, Maj., USAF
Deputy Director
Surveillance & Photonics Directorate

If your address has changed or if you wish to be removed from the Rome Laboratory mailing list, or if the addressee is no longer employed by your organization, please notify Rome Laboratory/OCSS, Rome, NY 13441. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document require that it be returned.

REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	August 1997	In-House, Jan 93 - Sep 93

4. TITLE AND SUBTITLE

A NEW EFFICIENT ALGORITHM FOR PDF APPROXIMATION

5. FUNDING NUMBERS

PE - 62702F
PR - 4506
TA - 11
WU- 0T

6. AUTHOR(S)

Lisa K. Slaski and Dr. Murali Rangaswamy

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Rome Laboratory/OCSS
26 Electronic Pky
Rome, NY 13441-4514

8. PERFORMING ORGANIZATION REPORT NUMBER

RL-TR-97-73

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Rome Laboratory/OCSS
26 Electronic Pky
Rome, NY 13441-4514

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

RL-TR-97-73

11. SUPPLEMENTARY NOTES

Rome Laboratory Project Engineer: E. Douglas Lynch/OCSA/315-330-4515

12a. DISTRIBUTION AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Classical radar signal processing techniques assume that the signal interference is Gaussian in nature. However, it has been shown that this interference or clutter is not always Gaussian. When non-Gaussian clutter exists, other signal processing techniques which are optimal, or more robust in non-Gaussian clutter may be more effective than the classical techniques. This requires determination of the clutter characteristics for each clutter region and then applying the appropriate signal processing technique to the data ideally in "real-time". In order to achieve "real-time" it is necessary to determine this approximate Probability Density Function (PDF) using small sample data set sizes. However, until the development of the Ozturk Algorithm, there has not existed an efficient algorithm to determine an approximate PDF for a small clutter data sample set. The Ozturk Algorithm is a new statistical algorithm capable of approximating the PDF of a set of random data using on the order of 100 sample points, whereas classical techniques typically require thousands of samples. It consists of two parts, a Goodness-of-fit Test and the PDF Approximation. The Goodness-of-fit Test determines whether a sample data set is statistically consistent with a given PDF. The PDF Approximation selects the "best" approximate PDF from a variety of PDFs and is simply an extension of the Goodness-of-fit Test. This report describes the Ozturk Algorithm and shows an application of the algorithm to some temporal L-band radar clutter data.

14. SUBJECT TERMS

probability & statistics, non-Gaussian clutter, clutter modeling, Ozturk Algorithm, radar

15. NUMBER OF PAGES

36

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

ABSTRACT

Classical radar signal processing techniques assume that the signal interference is Gaussian in nature. However, it has been shown that this interference or clutter is not always Gaussian. When non-Gaussian clutter exists, other signal processing techniques which are optimal or more robust in non-Gaussian clutter may be more effective than the classical techniques. This requires determination of the clutter characteristics for each clutter region and then applying the appropriate signal processing technique to the data ideally in 'real-time'. In order to achieve 'real-time' it is necessary to determine this approximate PDF using small sample data set sizes. However, until the development of the Ozturk Algorithm, there has not existed an efficient algorithm to determine an approximate PDF for a small clutter data sample set.

The Ozturk Algorithm is a new statistical algorithm capable of approximating the PDF of a set of random data using on the order of 100 sample points, whereas, classical techniques typically require thousands samples. It consists of two parts, a Goodness-of-fit Test and the PDF Approximation. The Goodness-of-fit Test determines whether a sample data set is statistically consistent with a given PDF. The PDF Approximation selects the 'best' approximate PDF from a variety of PDFs and is simply an extension of the Goodness-of-fit Test.

This report describes the Ozturk Algorithm and shows an application of the algorithm to some temporal L-band radar clutter data.

TABLE OF CONTENTS: page

1. Introduction to the Ozturk Algorithm:

The Ozturk Algorithm was developed by Dr. Aydin Ozturk, while he was a visiting Professor at Syracuse University, Syracuse, NY. The Ozturk Algorithm consists of two major parts, the Goodness-of-fit Test, and the PDF Approximation. The Goodness-of-fit Test determines whether or not a sample data set is statistically consistent with a given PDF. The PDF Approximation is an extension of the Goodness-of-fit Test, and results in the selection of a 'best' PDF which approximates the sample data set, using a closest linear distance measure.

This algorithm takes on the order of 100 data samples from any random data set to perform the Goodness-of-fit Test and PDF Approximation. It has been extensively tested for independent random data generated from a known distribution. It also, appears to work well with radar clutter data. The qualifier, 'appears' is used since the radar clutter data is of unknown distribution and thus it is more difficult to judge the result of the approximate PDF selection.

Note, that as the number of data samples used by the algorithm to determine the approximate PDF is increased (above a couple hundred samples), the algorithm becomes numerically inefficient and computationally intensive.

The objective of this report is to introduce the concept behind the Ozturk Algorithm without the overuse of detailed mathematics, which can make the explanation cumbersome and at times confusing. A more detailed discussion of the statistical mathematics used in the algorithm can be found in the references.

2. Advantages of the Ozturk Algorithm

Classical techniques for determining a 'good' PDF fit for a set of data require large data sets (on the order of 10,000 points). Also, the researcher must first select a PDF to test the data set against and use the appropriate test for this PDF. If the PDF fails to fit the data, then the researcher must select another PDF to test the data set against with another separate test for this new PDF. In other words, classical techniques provide an answer to the question "is a set of random data statistically consistent with a specified PDF ?".

The advantage of the Ozturk Algorithm is that only one test is performed, using a variety of PDF's on a much smaller data set (on the order of 100 points). The test determines which of the PDF's available best fits the data set. Furthermore, the algorithm provides a graphical representation of the goodness-of-fit and PDF approximation. Also, estimates of location, scale and shape parameters of the approximating PDF are available as outputs of the algorithm.

3. Detailed Algorithm Description:

This section describes the details of the algorithm and is organized by the two parts of the algorithm, 1) the Goodness-of fit Test and 2) the PDF Approximation.
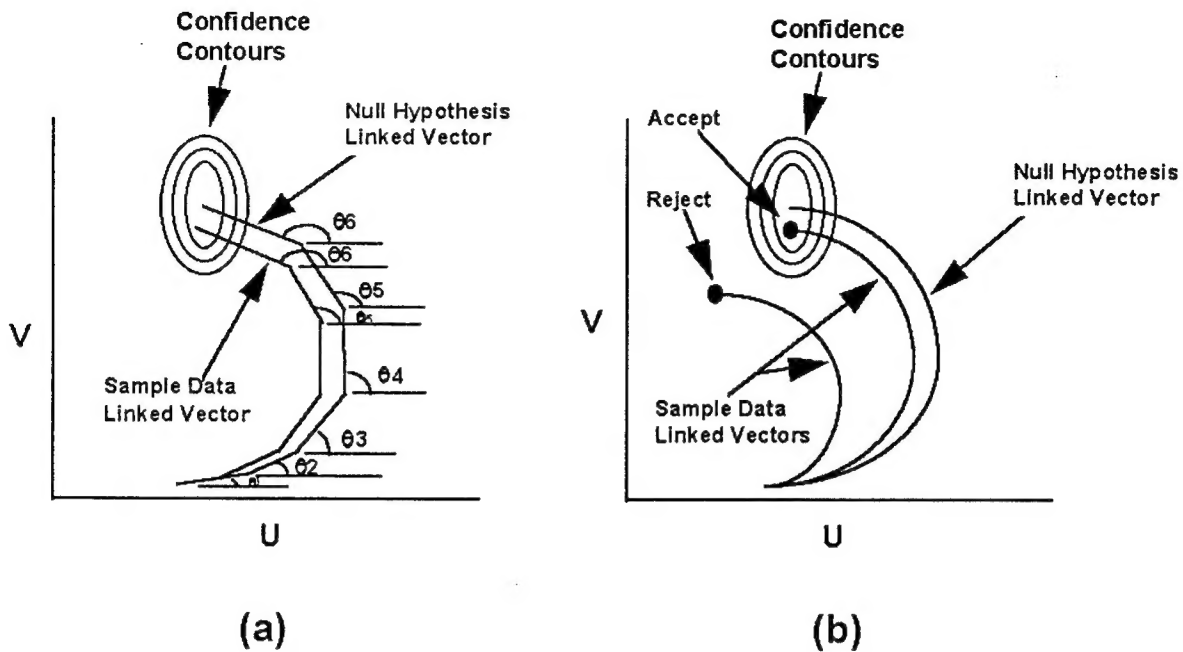
3.1. Goodness-of-fit Test:

The goodness-of-fit test is a complex algorithm which determines if the sample data provided to the algorithm is statistically consistent with a given distribution (the null hypothesis). Typically the sample data is tested against a standard Gaussian distribution. However, it may be tested against any available distribution.

In the Ozturk Algorithm, the reference distribution is the standard Gaussian distribution and the null hypothesis is the distribution against which the sample data is to be tested. Linked vectors are constructed for both the null hypothesis as well as the sample data set. The confidence contours are constructed around the terminal point of the null hypothesis linked vector.

3.1.1. The Linked Vector

The algorithm provides a graphical method for observing the consistency of the sample data against a null hypothesis by producing two loci of linked vectors and a set of confidence contours as shown in **Figure 1**.

Figure 1: Goodness-of-fit Test
a) N = 6 points   b) Large N

To obtain the linked vectors, first consider 1) the sample data set:

$$X_1, X_2, X_3, \ldots . X_N$$

with mean μ_x, standard deviation σ_x, and length N, and 2) a null hypothesis data set generated from any available distribution against which the sample set will be tested:

$$Z_1, Z_2, Z_3, \ldots . Z_N$$

with zero mean, unit variance and length N, and 3) a reference data set generated from the standard Gaussian:

$$W_1, W_2, W_3, \ldots . W_N$$

Next reorder all data sets (ordered statistics) with the smallest value first:

$$X_{1:N}, X_{2:N}, X_{3:N}, \ldots . X_{N:N}$$

$$Z_{1:N}, Z_{2:N}, Z_{3:N}, \ldots . Z_{N:N}$$

$$W_{1:N}, W_{2:N}, W_{3:N}, \ldots . W_{N:N}$$

Let, $y_{i:N}$, for the sample linked vector be defined as:

$$y_{i:N} = \frac{X_{i:N} - \mu_x}{\sigma_x}$$

The magnitude of the sample linked vector is the absolute value of $y_{i:N}$. Also let, $t_{i:N}$, for the null hypothesis be defined as the expected value of the ith ordered statistic of the null hypothesis distribution:

$$t_{i:N} = E[z_{i:N}] .$$

The magnitude of the null hypothesis linked vector is the absolute value of $t_{i:N}$. Finally let, $m_{i:N}$, be defined as the expected value of the ith ordered statistic of the reference distribution, the standard Gaussian:

$$m_{i:N} = E[w_{i:N}] .$$

The expected values are obtained through a Monte-Carlo simulation consisting of 2000 generated data sets for both the reference data

set and the null hypothesis.

Note that when the null hypothesis is the same as the reference distribution, then $t_{i:N}$ for the null hypothesis is simply:

$$t_{i:N} = m_{i:N} = \mathbf{E}[w_{i:N}] \, .$$

The set of angles associated with each linked vector is defined as:

$$\theta_i = \pi \phi (m_{i:N}) \, ,$$

where:

$$\phi (a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} \exp\left(-\frac{t^2}{2}\right) dt .$$

Next set up the co-ordinate system $Q_k = [u_k, v_k]$, where:

$$u_k = \frac{1}{k} \sum_{i=1}^{k} |y_{i:N}| \cos\theta_i \qquad ; \; k=1,2,3,\ldots N$$

$$v_k = \frac{1}{k} \sum_{i=1}^{k} |y_{i:N}| \sin\theta_i \qquad ; \; k=1,2,3,\ldots N$$

for the sample linked vector. The null hypothesis linked vector is obtained by replacing $y_{i:N}$ with $t_{i:N}$ in u_k and v_k above.

Note that the angle theta is solely dependent on the reference distribution for all linked vectors, while the magnitude is solely dependent on the data chosen for the linked vector (e.g. for the sample linked vector, the magnitude is dependent on the normalized

ordered statistic of the sample data, for the null hypothesis linked vector, the magnitude is dependent on the expected value of the ordered statistic of the monte-carlo simulation of the null distribution).

Further note that $y_{i:N}$ and $t_{i:N}$ are ordered statistics from smallest to largest, while the magnitudes of $y_{i:N}$ and $t_{i:N}$, may no longer be true ordered statistics, due to standardization. If $y_{i:N}$ and $t_{i:N}$ contain negative values due to standardization, then their magnitudes would begin large, decrease to approximately zero and then increase again.

Also, when N, the length of the data set, is large (on the order of 50 points), then the linked vector is a smooth arc.

3.1.2. The Confidence Contours

The algorithm provides quantitative information as to how consistent the sample data set is with the null hypothesis distribution by the use of confidence contours. These contours are shown graphically in **Figure 1**. If the end point of the sample data set falls within one or more of these contours, then the sample data is considered to be statistically consistent with the null hypothesis with a given confidence level based on the confidence contour. Also note, that if the sample data is truly consistent with the null hypothesis, then the trajectory of the sample linked vector is likely to follow that of the null hypothesis linked vector.

3.1.2.1 Basic Concept

Consider first that the linked vector for the null hypothesis is based on the expected values of the order statistic z for 2000 monte-carlo simulations. Thus if one considers just one point along the linked vector, in particular the end point, the monte-carlo simulation provides 2000 points of which only the expected value is plotted. However, these 2000 points can also be analyzed for their distribution.

To determine the confidence contours for the null hypothesis, fit a three dimensional bell shape (bivariate Gaussian) curve to the 2000 points arising from the distribution of the (Monte-Carlo) end points for the null hypothesis linked vector. Then plot the contours of constant density of this distribution for various values of the parameter alpha, (e.g., 0.01, 0.05, and 0.10), where alpha is the probability that the end point falls outside the specified contour given that the data is from the null hypothesis distribution. Then unity minus alpha is known as the confidence level and the corresponding contour is known as the confidence contour. Alpha is known as the significance level of the test.

This may be repeated for any of the N points of the ordered statistic, z, along the null hypothesis linked vector. If the sample data is truly consistent with the null hypothesis, then the sample data's linked vector trajectory is will pass through a series of hoops defined by the confidence contours from all points along the null hypothesis linked vector and end within the last set of confidence contours. However, it is not necessary to clutter up the graphics with all these confidence contours, as the human eye can readily detect whether or not the linked vectors are closely following the same trajectory. Thus, only the last set of confidence contours are typically provided.

As the significance level of the test increases, the corresponding confidence level decreases and the confidence contours decrease in size. The closer the end point of the linked vector for the sample data falls to the center of the confidence contours, the more likely it is that the sample is from the null hypothesis.

Also, for a given sample size, N, note that the ith angle which is dependent solely on the reference distribution remains unchanged and is used by all linked vectors. Also, the magnitude of the sample data linked vector is solely dependent on the sample data set. Thus, the linked vector for the null hypothesis distribution and the theta values associated with the sample size, may be tabulated based on N (and 2000 monte-carlo simulations). This table, which is dependent on N, for a given null hypothesis, and the theta values dependent on N, may be stored and recalled when desired. This can significantly reduce the computation requirements of the algorithm for a 'real-time' application.

3.1.2.2 Detailed Description

As previously described, the confidence contours are contours of the probability distribution of the null hypothesis linked vector end points from a monte-carlo simulation (2000 points). To analytically determine the confidence contours, the joint PDF of u_i and v_i must be known. However, this joint PDF is difficult to determine analytically. Thus, it is necessary to rely upon empirical results.

The central limit theorem states that if M is sufficiently large and the random variables x_k, where $k=[0,1,2,3.... M]$, are independent, and identically distributed, then under general conditions, the density function of their sum, properly normalized, tends to a normal curve as M approaches infinity. Assuming the conditions are satisfied, the central limit theorem allows the researcher to approximate the marginal PDFs of u_i and v_i as Gaussian for the monte-carlo simulation. In addition, through empirical analysis, it has been observed that the joint PDF of u_i and v_i can often be approximated as bivariate Gaussian. The bivariate Gaussian PDF is defined as:

$$f_{u,v}(u_i, v_i) = \frac{\exp\left\{\frac{-1}{(2(1-\rho^2))}\left[\frac{(u_i-\mu_u)^2}{\sigma_u^2} - \frac{(2\rho(u_i-\mu_u)(v_i-\mu_v)}{\sigma_u\sigma_v} + \frac{(v_i-\mu_v)^2}{\sigma_v^2}\right]\right\}}{(2\pi\sigma_u\sigma_v\sqrt{1-\rho^2})}$$

where:

$$\mu_u = E(u_i), \quad \mu_v = E(v_i)$$
$$\sigma_u^2 = Var(u_i), \quad \sigma_v^2 = Var(v_i)$$
$$\rho = E\left[\frac{(u_i-\mu_u)(v_i-\mu_v)}{\sigma_u\sigma_v}\right]$$

The mean, variance and correlation coefficient are all obtained empirically.

It is well known that the locus of constant values of this PDF

will be an ellipse, and that it is maximum at the point, (μ_u, μ_v). The ellipse degenerates into a circle when the variance of u_i and v_i are equal and the correlation coefficient is zero. Also, it degenerates into a line that passes through its maximum as the correlation coefficient approaches ± 1.

In order to maximize the likelihood that the random variables u_i and v_i are Gaussian, the number of monte-carlo simulations should be large. To reduce the number of required simulations, the Ozturk Algorithm incorporates the Johnson System of Transformation (reference #3). This technique transforms limited data sets in such a way as to approximate the resulting Gaussian distribution. Thus, only 2000 monte-carlo simulations are required to determine the appropriate marginal Gaussian PDFs and thus the bivariate PDF and its associated confidence contours.

3.2. PDF Approximation

To select the 'best' approximate PDF the algorithm develops the PDF Approximation Chart as a visual aid. This chart is simply an extension of the Goodness-of-fit Test. In the Goodness-of-fit Test, a sample data set is tested for statistical consistency against a null hypothesis of a selected distribution. The PDF

Approximation Chart takes this a step further by providing other distributions. These distributions are computed in the same manner as described previously for the null hypothesis in that the magnitude of the linked vector is computed from the expected value of the ordered statistic of 2000 monte-carlo simulations. However, the angle theta is still computed from the reference distribution and confidence contours are computed only for the null hypothesis.

Refer to **Figure 2** as an example of a PDF Approximation Chart. If all the linked vectors for the these various distributions were provided in the graphics, the plot would soon become too cluttered to properly interpret the data. Also, the primary information from the linked vectors is contained in the location of their respective end points. Therefore, only the end points of all linked vectors are provided in the approximation chart, along with the confidence contours (not shown in **Figure 2**) for the selected distribution (null hypothesis).

For distributions dependent only on mean and variance (no shape parameters), such as Gaussian, there exists only one unique linked vector and thus only one point on the approximation chart is plotted.

For distributions dependent on a single shape parameter, such as Weibull, different values of the shape parameter result in

V

U
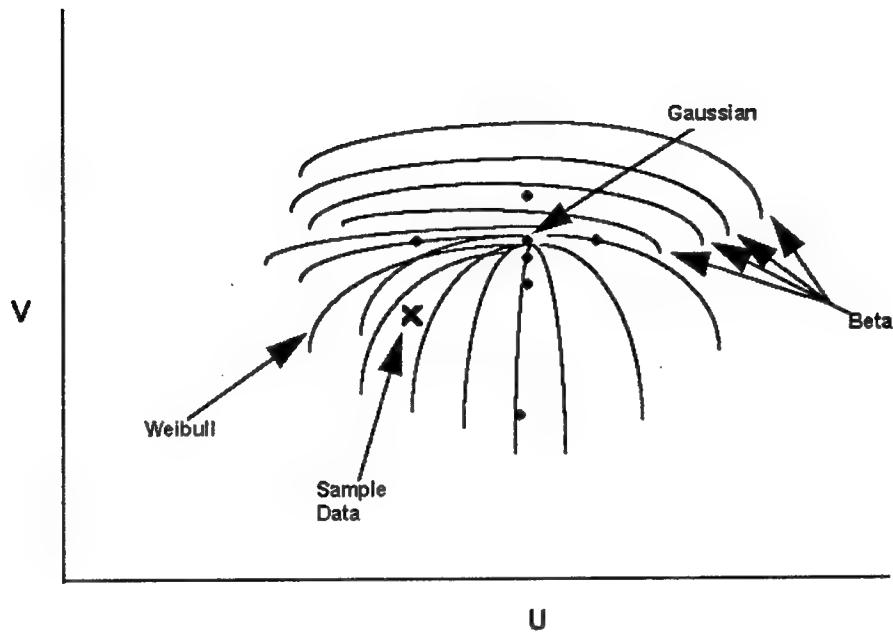
Figure 2: PDF Approximation Chart

different linked vectors. Consequently, the end point of the
linked vectors is also dependent on the shape parameter. The end
points corresponding to different shape parameter values are joined
to obtain a single curve on the Approximation Chart. This curve
provides a unique representation for the PDF dependent on a single
shape parameter.

Similarly, for a distribution dependent on two shape parameters,

page 13

such as the Beta distribution, a series of linked vectors must be computed in order to plot the surface on which the end point travels for varying shape parameters. This is performed by holding the first shape parameter constant and varying the second shape parameter to generate a curve, then changing the first shape parameter and again hold it constant while varying the second shape parameter, etc... until a family of curves is produced over the surface that the distribution occupies.

Thus an approximation chart such as that shown in **Figure 2** can be produced. This chart can then be used to identify the distribution that best approximates the sample data.

If there is some reason to believe apriori that the sample data comes from a given distribution, e.g. Gaussian or Weibull with a given shape parameter, then the chosen null hypothesis would be selected to be this distribution. The goodness-of-fit test provides information as to whether or not the sample data is statistically consistent with the selected null hypothesis, a portion of this information is still present in the approximation chart in the constructed confidence contours. If the end point of the linked vector for the sample data falls within the confidence contours of the chosen null hypothesis, then no further work is required, as the sample data is statistically consistent with this hypothesis. However, if the end point for the sample falls outside

of the confidence contours then it is not statistically consistent with the null hypothesis. To select the best approximate PDF, the algorithm chooses the closest distribution to the sample and estimates the shape parameters of this distribution if required. The algorithm can also provide a rank order of selection for PDF approximation of all available distributions based on their respective distances from the sample data.

The important point to note here is that although the linear distance from the sample data is the criteria used to order the various distributions, this is not the most accurate method, although it is the simplest. The circular confidence contours, as shown in **Figures 1** is a special case of the confidence contours. In general, the confidence contours have been found to be elliptical and nearly circular for larger values of N (on the order of 50-100 points). However, under some conditions they may become quite elongated in shape. Therefore, the linear distance can give an idea of which PDF is the best approximating PDF, but the only accurate method is by determining the significance level for each PDF. However, due to the complexity and numerical computation efficiency required, this is not as simple as it seems.

4. L-Band Radar Application

The following is an example of the analysis capability of Ozturk's algorithm using actual radar measurement data from an L-band ground radar located in the RL/OC Surveillance Facility.

The radar data used here is a time sequence from a single range cell in a 'stare' or 'searchlight' mode which contains a strong clutter signal. It has been shown that clutter data from a single range cell is locally Gaussian. This means that the individual quadrature components of the received signal are Gaussian. Thus the clutter magnitude is expected to be Rayleigh.

Figure 3 shows the raw radar data from a single PRI (containing many range cells at a given azimuth direction). From this data range cell (bin) 90 was chosen as the test cell for strong clutter.

Figure 4 shows the magnitude of the received radar signal from a single range cell for approximately 1000 PRIs. The cell chosen was bin 90. As can be seen and as expected for a ground radar the data is highly correlated with a fluctuating signal riding on top of a slowly varying deterministic signal. The deterministic portion of the signal can be thought of as the return from the ground, buildings and other stationary objects within the range cell. The fluctuating portion of the signal is the clutter of interest. This
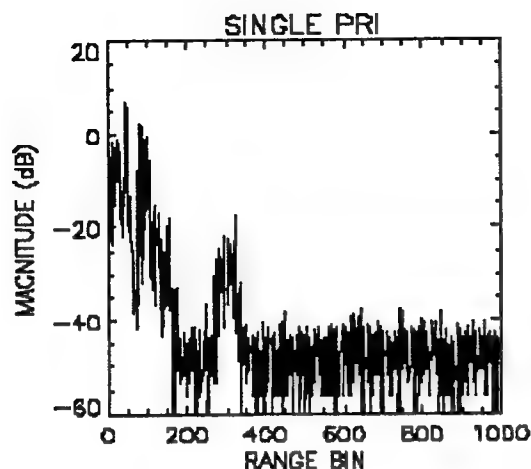
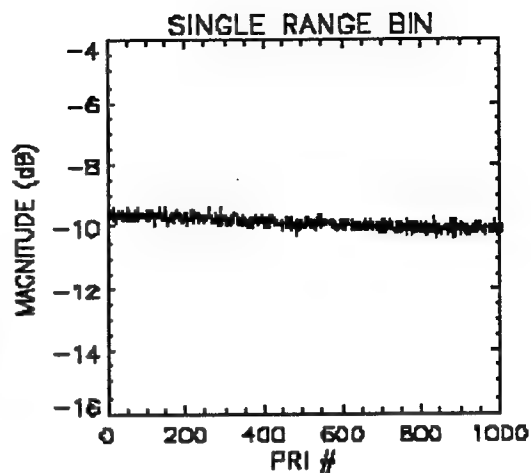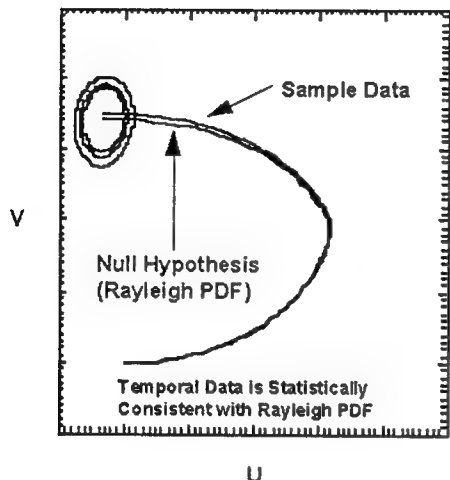Figure 3: L-Band Radar Data   Figure 4: L-Band Radar Data
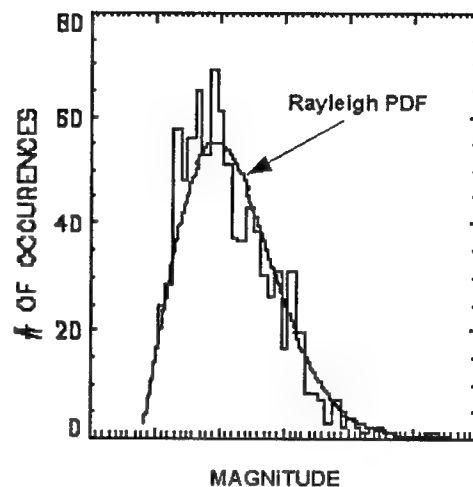   (temporal data)

would consist of signal returns from such objects as blowing grass, tree limbs and other non-stationary objects. Thus, in order to analyze the clutter of interest, a two-pulse cancellation is performed on the data. Note that this will provide uncorrelated data, but not necessarily independent data for the Ozturk Algorithm. After cancellation, the data was found to be 10dB higher than the noise, and thus contains a clutter signal.

Figure 5 shows the Goodness-of-fit Test for the first 100 PRIs of the two pulse cancelled radar data against a Rayleigh PDF (the null hypothesis). This test shows that the 100 radar data points are statistically consistent with the Rayleigh PDF as expected. As can be seen from **Figure 6**, the histogram of the two-pulse cancelled radar data fits the Rayleigh PDF fairly well.

Figure 5: L-Band Radar Data
Goodness-of-fit Test

Figure 6: L-Band Radar Data
1000 point Histogram

As noted, temporal radar clutter data from a ground radar is typically expected to be locally Gaussian. However, for certain cases, especially for spatial radar clutter data, it has been shown that the clutter is not necessarily Gaussian. The algorithm, has been tested extensively with theoretical data and limitedly with temporal radar data. As yet it has not been tested with spatial radar data, which is expected to be non-Gaussian in general. All cases to date, of the limited testing with temporal radar data, has shown that the clutter is statistically consistent with the Rayleigh PDF.

In order to show the application of the Ozturk Algorithm for PDF

approximation, a theoretical data sample is generated. In this
case 1000 sample data points have been generated from the lognormal
distribution. A shape parameter of 0.8 was chosen such that any
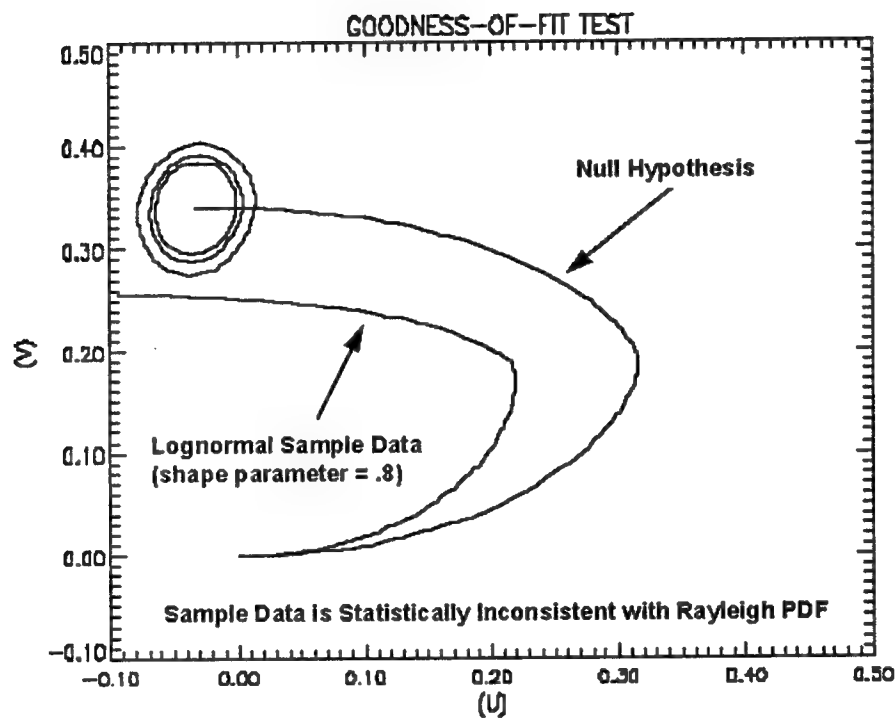100 point data sample set from this distribution will likely be

Figure 7: Lognormal Data

statistically inconsistent with the Rayleigh PDF. As shown in
Figure 7, the Ozturk Algorithm has determined that this sample set

is statistically inconsistent as expected.

Thus, since the data is statistically inconsistent, the second portion of the algorithm is exercised in order to obtain an approximate PDF. From **Figure 8**, it can be seen that the data sample linked vector's end point lies closest to the Type-2 Gumbel
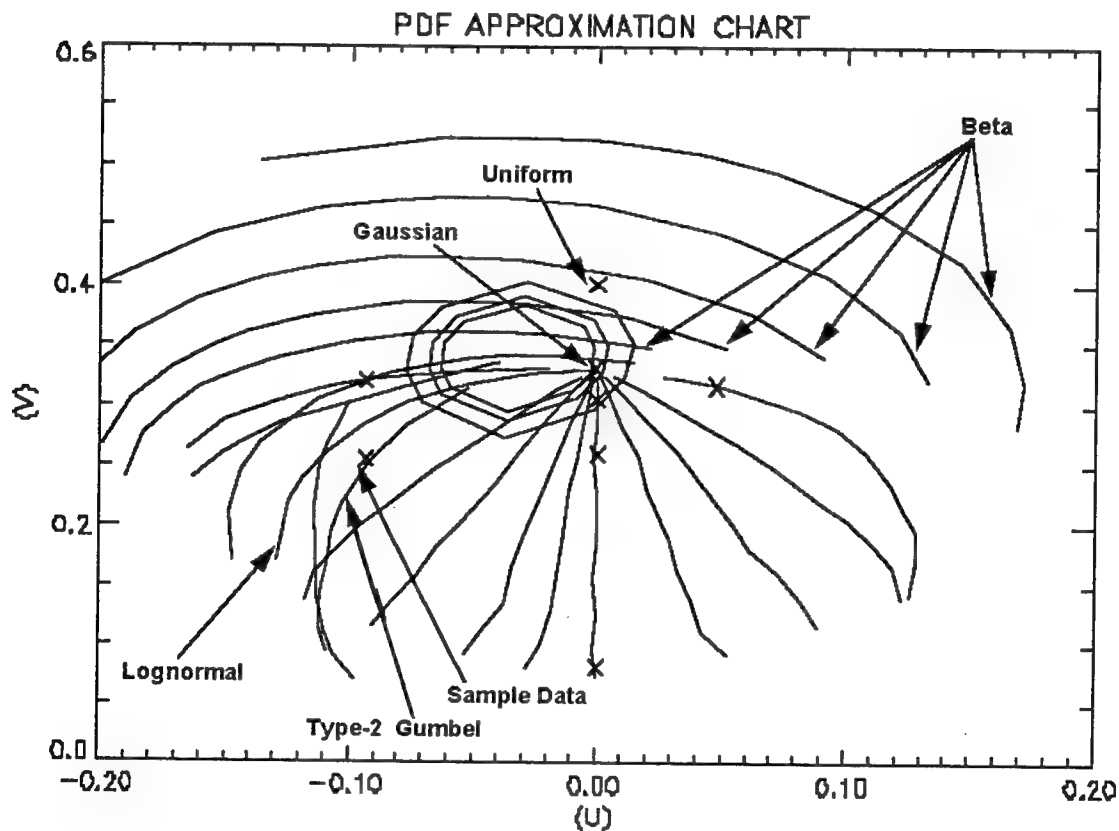

Figure 8: Lognormal Data is Approximated by Gumbel PDF

PDF. Thus, this PDF is chosen and its shape parameter is estimated to provide the approximate PDF. Note that the algorithm does not identify the true PDF, but rather, approximates the true PDF with

page 20

one it selects as the 'best' approximate PDF.

In **Figures 9 and 10** the histrogram of the data used for determining the approximate PDF is overlaid with the Rayleigh, Type-2 Gumbel and Lognormal PDFs. They show, in general, that the data definitely does not fit the Rayleigh PDF and that the Type-2 Gumbel PDF is very similar in shape to the Lognormal PDF. However, a 100 point histogram is not a very good histogram, in the sense that there just isn't enough data to appreciate the shape of the histogram. Also, changing the bin size of the a 100 point histogram can have profound effects on the shape of the histogram. Thus the histogram and associated PDFs were extended to the 1000 points available.

Figures 11 and 12 show the 1000 point histograms overlaid with the Rayleigh, Type-2 Gumbel and Lognormal PDFs. Again, the histogram of the data obviously shows that the data is not consistent with the Rayleigh PDF. Also, the Type-2 Gumbel, determined from the first 100 points of the sample set, still approximates the Lognormal rather well. Furthermore, it could even be said that the Type-2 Gumbel PDF approximates this particular set of 1000 points better than the distribution from which they were generated, since the Gumbel PDF seems to fit the histogram better than the Lognormal PDF.

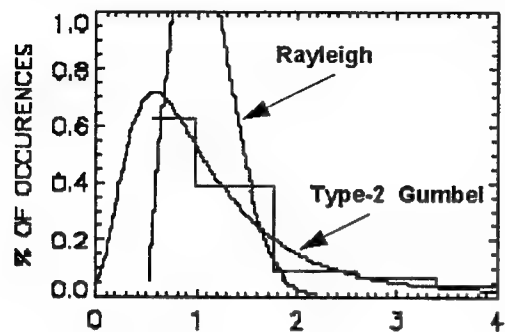**Figure 9: 100 Samples from the Lognormal PDF**

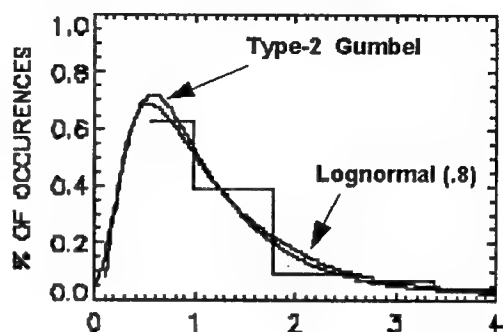**Figure 10: 100 Samples from the Lognormal PDF**

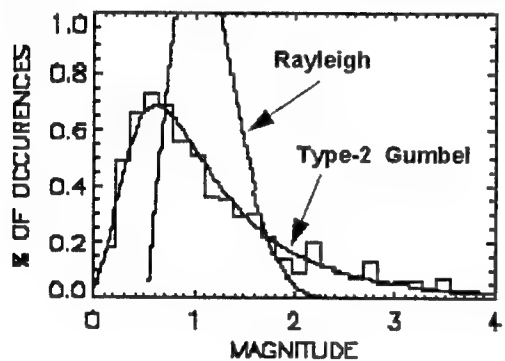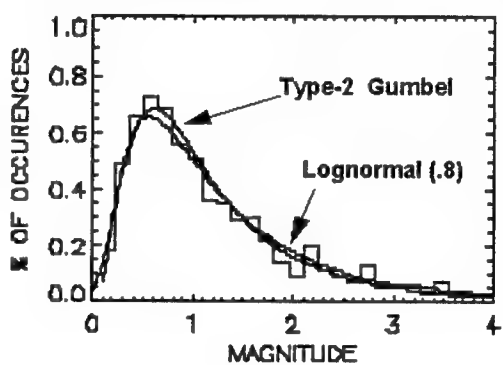**Figure 11: 1000 Samples from the Lognormal PDF**

Figure 12: 1000 Samples from the Lognormal PDF

5.0 Conclusions

The Ozturk Algorithm seems to perform as advertized. It efficiently determines an approximate PDF to a set of random data using on the order of 100 sample data points. It can select an approximate PDF from a variety of PDFs. In theory as well as in practice, the algorithm has so far performed well.

The most difficult assumption required by the algorithm for the radar engineer to deal with is that the sample data must be independent. Radar clutter data, in general, is not necessarily independent. Although obtaining uncorrelated samples may not be terribly difficult, insuring independence is very difficult for the radar engineer. Data samples which are uncorrelated are not necessarily independent. The effect of uncorrelated, but not necessarily independent data on the algorithm's performance is a relevant issue. However, so far, the algorithm has appeared to perform well with radar measurement data.

This promising algorithm needs to be exercised more fully to understand its advantages and application in the radar engineering field. One of the greatest potentials of this algorithm is providing the capability of determining the clutter characteristics using only a relatively small data sample size. In a clutter environment containing non-Gaussian as well as Gaussian clutter,

this capability might allow for the radar engineer through other algorithms, generally referred to as expert systems, to select the appropriate signal processing technique for the given clutter region in order to obtain better target detection than the classical techniques which assume Gaussian clutter statistics.

Although the application described here is a radar engineering application, the Ozturk Algorithm may be very useful in many technical fields, both commercial and military.   It may be especially useful in any field which makes use of imaging techniques and detection algorithms.

6. References

1. "Signal Detection in Correlated Gaussian and Non-Gaussian Radar Clutter", M. Rangaswamy, P. Chakravarthe, Dr. D. Weiner, Dr. L. Cai, Dr. H. Wang, Dr. A. Ozturk, Rome Lab contract final report, RL-TR-93-79, Chapt 6, 1993.

2. "A New Statistical Goodness-of-fit Test Based on Graphical Representation", Aydin Ozturk and Edward J. Dudewicz, Technical Report no. 52, Department of Mathematics, Syracuse University, Syracuse, NY (1990).

3. "Multivariate Statistical Simulation", Mark Johnson, Wiley and Sons Inc. copyright 1987

4. "A General Algorithm for Univariate and Multivariate Goodness of Fit Tests Based on Graphical Representation", Aydin Ozturk, Communications in Statistics: Theory and Methods, Vol 20, No. 10, pages 3111-3138, 1991.

5. "A New method for Assessing Multivariate Normality with Graphical Applications", Aydin Ozturk and J. Romeu, Communications in Statistics: Theory and Methods, Vol 21, No. 1, pages 15-34, 1991.

MISSION
OF
ROME LABORATORY

Mission. The mission of Rome Laboratory is to advance the science and technologies of command, control, communications and intelligence and to transition them into systems to meet customer needs. To achieve this, Rome Lab:

    a. Conducts vigorous research, development and test programs in all applicable technologies;

    b. Transitions technology to current and future systems to improve operational capability, readiness, and supportability;

    c. Provides a full range of technical support to Air Force Material Command product centers and other Air Force organizations;

    d. Promotes transfer of technology to the private sector;

    e. Maintains leading edge technological expertise in the areas of surveillance, communications, command and control, intelligence, reliability science, electro-magnetic technology, photonics, signal processing, and computational science.

The thrust areas of technical competence include: Surveillance, Communications, Command and Control, Intelligence, Signal Processing, Computer Science and Technology, Electromagnetic Technology, Photonics and Reliability Sciences.